

none

none

none

© EPODOC / EPO

PN - JP2000305579 A 20001102  
PD - 2000-11-02  
PR - JP19990115713 19990423  
OPD - 1999-04-23  
TI - SPEECH DETECTING DEVICE  
IN - TERASAWA AKIRA; FUKUSHIMA MINORU; AKEYAMA HIROAKI  
PA - MATSUSHITA ELECTRIC WORKS LTD  
IC - G10L11/02 ; G10L21/02

© WPI / DERWENT

TI - Audio detector for intercoms used at residence, office, updates assessed value for noise power estimation in noise power estimation unit based on judgment of reference signal to be an audio signal  
PR - JP19990115713 19990423  
PN - JP2000305579 A 20001102 DW200138 G10L11/02 008pp  
PA - (MATW ) MATSUSHITA ELECTRIC WORKS LTD  
IC - G10L11/02 ;G10L21/02  
AB - JP2000305579 NOVELTY - A reference signal is extracted from communication channel and power of audio component and noise component are estimated in power and noise estimation units 1,2, respectively. Based on the estimated noise power, reference signal is judged to be audio signal. Assessed value for noise power estimation in noise power estimation unit is updated, based on judgment result.  
- USE - For intercoms; telephones used in residence, office, factory.  
- ADVANTAGE - By updating assessed value for noise power estimation, the background noise power estimation process is suspended when acoustic coupling component included in reference signal is large, and hence presence of audio component in reference signal can be accurately detected.  
- DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of amplifying call machine.  
- Noise estimation units 1,2  
- (Dwg.1/5)  
OPD - 1999-04-23  
AN - 2001-358627 [38]

© PAJ / JPO

PN - JP2000305579 A 20001102

none

none

none

none

none

none

- PD - 2000-11-02  
AP - JP19990115713 19990423  
IN - FUKUSHIMA MINORU; AKEYAMA HIROAKI; TERASAWA AKIRA  
PA - MATSUSHITA ELECTRIC WORKS LTD  
TI - SPEECH DETECTING DEVICE  
AB - PROBLEM TO BE SOLVED: To precisely detect whether a reference signal contains a speech signal or not in a system presenting a large acoustic coupling gain or under a condition with a high background noise level.  
- SOLUTION: A background noise power estimating and switching part 4 stops an estimation processing in a background noise power estimating part 2 when the ratio of an acoustic coupling gain component contained in a reference signal  $V_x$  is large. This makes a background noise power estimation value  $P_n$  becomes close to a value approximating a background noise power around a microphone 10 regardless of the phone call state. Thus, this device reduces a deterioration of a speech detection performance caused by a turn-around of a background noise on a far side to the microphone 10 in a system with the microphone 10 and a speaker 11 in a short distance causing a large acoustic coupling gain, and precisely detects whether a reference signal contains a speech signal or not under a condition with a high background noise level.  
- G10L11/02 ;G10L21/02

none

none

none

\* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

## DETAILED DESCRIPTION

---

### [Detailed Description of the Invention]

#### [0001]

[The technical field to which invention belongs] this invention relates to the voice detector for carrying a normal-mode-rejection function, a voice change function, etc. in the speaking circuit in the \*\*\*\* telephone call equipments (an interphone, telephone, PHS, etc.) used at a residence, an office, works, etc.

#### [0002]

[Description of the Prior Art] Generally, a voice detector is used in order to detect whether the acoustic signal collected with the microphone contains the sound signal. The typical example of composition of such a voice detector is shown in drawing 5. This voice detector VD' is equipped with the instant power presumption section 20, the background-noise power presumption section 21, and the comparison-test section 22. An integrating circuit or a digital filter with the property with loose falling that a standup is steep and etc. is realized, and the instant power presumption section 20 presumes the short-time average power of a reference signal (acoustic signal collected with a microphone). Moreover, an integrating circuit or a digital filter with a property with falling steep gently [ the background-noise power presumption section 21 / a standup ] etc. is realized, and the background-noise (background noise) level which exists regularly in a reference signal is presumed. Furthermore, by comparing with a predetermined threshold the ratio of the instant power estimate calculated by the instant power presumption section 20, and the background-noise power estimate calculated by the background-noise power presumption section 21, the comparison-test section 22 judges whether the reference signal contains the sound signal (detection), and outputs the binary signal (detecting signal) of H or L.

#### [0003]

[Problem(s) to be Solved by the Invention] However, when preparing above voice detector VD' in the internal circuitry of a \*\*\*\* telephone call machine (not shown) which has a loudspeaker and a microphone, the wraparound component (acoustic turnover component) from a loudspeaker is contained in the acoustic signal collected with a microphone. When the rate of the wraparound component contained in this acoustic signal is large, it becomes difficult to detect whether the speaker who is near the microphone which is the original purpose uttered voice. For example, the background-noise level near the telephone call terminal by the side of a far edge is large, and when a microphone collects the background noise by the side of a far edge through an acoustic turnover, the background-noise power estimate calculated by the background-noise power presumption section 21 in the above-mentioned conventional example becomes large. Consequently, also in the state where the speaker who is near the microphone uttered voice, the ratio of instant power estimate and background-noise power estimate is small, and there is a possibility that it may be incorrect-detected as not being a sound signal (non-voice) in the comparison-test section 22, without the ability exceeding a predetermined threshold.

[0004] Succeeding in this invention in view of the above-mentioned problem, the place made into the purpose is to offer the voice detector which can detect with a sufficient precision whether the sound signal is contained in the reference signal under the situation that background-noise level is large.

#### [0005]

[Means for Solving the Problem] It is used for the above-mentioned \*\*\*\* telephone call terminal of

the \*\*\*\* telephone call system which the \*\*\*\* telephone call terminal which has a microphone and a loudspeaker is connected to other telephone call terminals or \*\*\*\* telephone call terminals, and performs a half-duplex telephone call in order that invention of a claim 1 may attain the above-mentioned purpose. The instant power presumption section which presumes the instant power of the reference signal which is the voice detector which detects whether the signal transmitted to a channel is a sound signal, or it is a non-voice signal, and was taken out from the above-mentioned channel, The background-noise power presumption section which presumes the power of the background-noise component contained in the above-mentioned reference signal, While judging whether a reference signal is a sound signal or it is a non-voice signal based on the background-noise power estimate presumed in the instant power estimate and the above-mentioned background-noise power presumption section which are presumed in the above-mentioned instant power presumption section The 1st voice / non-voice judging section which holds the last judgment result until a judgment result is updated, It is characterized by having the background-noise power presumption change section which changes updating/halt of the background-noise power estimate in the above-mentioned background-noise power presumption section. When the rate of the acoustic turnover component contained in a reference signal is large, while suspending processing of the background-noise power presumption section by the background-noise power presumption change section Since the judgment by the 1st voice / non-voice judging section is performed based on the background-noise power estimate which asked in the situation that the rate of the acoustic turnover component contained in a reference signal is small, and the background-noise power presumption section held. When the background noise in the telephone call terminal by the side of a far edge is sent out from a loudspeaker and turns to a microphone The situation that it becomes impossible to detect a sound signal even if the ratio of the voice component which the near end side speaker in the acoustic signal which a microphone collects emits, and the other background-noise component originates in a bird clapper small and the sound signal is contained in the reference signal can be reduced. It is detectable with a sufficient precision whether a reference signal is a sound signal under the situation that background-noise level is large.

[0006] voice the non-detecting section where a reference signal finds voice the non-detecting duration detected as it is a non-voice signal according to the judgment result according [ invention of a claim 2 / on invention of a claim 1, and ] to the voice / non-voice judging section of the above 1st -- a time check -- with the section It has the 2nd voice / non-voice judging section which judges any of a sound signal and a non-voice signal the above-mentioned reference signals are from voice the non-detecting duration found by the section. this voice non-detecting section -- a time check -- Voice the non-detecting duration found by the section is abbreviation regularity over time to be the phoneme duration grade of human being's voice. this -- the 2nd voice / non-voice judging section -- the above-mentioned voice non-detecting section -- a time check -- And it is characterized by judging all the reference signals of this voice non-detecting duration to be a sound signal, and changing, when the above-mentioned voice non-detecting duration is the pitch period grade of human being's voice. [ when a sound signal is not detected in the 1st voice / non-voice judging section under the situation that background-noise level is very large ] voice the non-detecting section -- a time check -- almost uniformly, in being almost equal to the pitch interval of human being's voice, while voice the non-detecting duration measured in the section is an audio phoneme duration grade Since a reference signal is anew judged as a sound signal in the 2nd voice / non-voice judging section, in a \*\*\*\* telephone call system with large background-noise level, a sound signal can be detected with a still more sufficient precision.

[0007]

*Geir Ichinoya  
Olivet*  
[Embodiments of the Invention] (Operation gestalt 1) Drawing 1 is the block diagram showing the \*\*\*\* telephone call machine M which has the voice detector VD1 in the operation gestalt 1 of this invention. This \*\*\*\* telephone call machine M is equipped with a microphone 10, a loudspeaker 11, the microphone amplifier 15, the loudspeaker amplifier 16, the voice detector VD1, and voice switch VS, and is connected with other \*\*\*\* telephone call machines etc. through a circuit. Voice switch VS here The acoustic turnover from a loudspeaker 11 to a microphone 10, And it is what oppresses a howling by reducing the gain of the closed loop formed of the wraparound by the side of a circuit. The transmission side attenuator 12 inserted on the transmission signal line for transmitting the

*echo*

sound signal (transmission signal) which collects a sound with a microphone 10 to a circuit. It has the receiver side attenuator 13 inserted on the receiver signal line for transmitting the sound signal (receiver signal) which received from the circuit to a loudspeaker 11, and the amount control section 14 of insertion losses which controls the gain of the transmission side attenuator 12 and the receiver side attenuator 13 according to a talk state. It \*\*, and in the amount control section 14 of insertion losses, a transmission-and-reception talk signal is observed, a talk state is judged, and the gain of the transmission side attenuator 12 and the gain of the receiver side attenuator 13 are appropriately set up according to a talk state.

[0008] The instant power presumption section 1 which presumes the instant power of the reference signal (transmission signal)  $V_x$  which took out the voice detector VD1 concerning this invention from the channel (transmission signal line), The background-noise power presumption section 2 which presumes the power of the background-noise component contained in the reference signal  $V_x$ , While judging whether the reference signal  $V_x$  is a sound signal or it is a non-voice signal based on the background-noise power estimate  $P_n$  presumed in the instant power estimate  $P_s$  and the background-noise power presumption section 2 which are presumed in the instant power presumption section 1 It has the 1st voice / non-voice judging section 3 holding the last judgment result, and the background-noise power presumption change section 4 which changes updating/halt of the background-noise power estimate  $P_n$  in the background-noise power presumption section 2 until a judgment result is updated.

[0009] It is constituted by an integrating circuit or a digital filter etc. in which falling has a loose property steeply [ a standup ] as for the instant power presumption section 1.. Moreover, it is constituted by the integrating circuit or digital filter in which falling has a steep property gently [ a standup ] as for the background-noise power presumption section 2.

[0010] The comparator CP 1 which, on the other hand, outputs the binary signal D1 of H or L for the instant power estimate  $P_s$  outputted from the instant power presumption section 1 as the 1st voice / non-voice judging section 3 are shown in drawing 2 as compared with the predetermined threshold  $P_{s0}$  the ratio of the instant power estimate  $P_s$  and the background-noise power estimate  $P_n$  outputted from the background-noise power presumption section 2 -- with divider 3a which calculates  $P_s/P_n$  It is constituted by the comparator CP 2 which outputs the binary signal D2 of H or L for output-value  $P_s/P_n$  of divider 3a as compared with predetermined threshold delta; and AND-operation section 3b which searches for the AND of two binary signals D1 and D2. The instant power estimate  $P_s$  is larger than a threshold  $P_{s0}$  ( $P_s > P_{s0}$ ), and in this operation gestalt, it \*\*, when output  $P_s/P_n$  of divider 3a is larger than threshold delta ( $P_s/P_n > \text{delta}$ ), it judges with a sound signal, and in the case of others, it judges with a non-voice signal. Here, a threshold  $P_{s0}$  is a threshold which specifies the minimum level of a sound signal, and threshold delta is a threshold which specifies the minimum ratio of sound signal level and back-ground-noise level.

[0011] The background-noise power presumption change section 4 is turned on and off by the control signal  $V_s$  outputted from the amount control section 14 of insertion losses of voice switch VS, and the input of the reference signal  $V_x$  over the background-noise power presumption section 2 consists of close / a switch which carries out OFF. And the background-noise power presumption section 2 serves as the update mode, when the background-noise power presumption change section 4 turns on and the reference signal  $V_x$  is inputted, and when the background-noise power presumption change section 4 turns off and the reference signal  $V_x$  is not inputted, it serves as halt mode. In the update mode, the background-noise power presumption section 2 updates the background-noise power estimate  $P_n$  serially with reference to the reference signal  $V_x$  here. Moreover, in halt mode, the background-noise power presumption section 2 suspends the above-mentioned data processing, and holds the value calculated before it as background-noise power estimate  $P_n$ . halt mode

[0012] Here, when a talk state was judged to be a receiver state, while the amount control section 14 of insertion losses of voice switch VS turned off the background-noise power presumption change section 4 with the control signal  $V_s$ , when it judges with a transmission state, it turns on the background-noise power presumption change section 4 with a control signal  $V_s$ . Since it \*\*, it becomes halt mode in a receiver state in the background-noise power presumption section 2 and it becomes the update mode in a transmission state, it sets to the voice detector VD1. When the rate of

the acoustic turnover component contained in the reference signal Vx is large, by suspending presumed processing of the background-noise power presumption section 2, it can consider as the value which approximated the background-noise power of the microphone 10 circumference, without depending the background-noise power estimate Pn on a talk state. Consequently, the distance between a microphone 10 and a loudspeaker 11 is short, and degradation of the voice detectability ability by the background noise by the side of a far edge turning to a microphone 10 also in a system with large acoustic turnover gain can be reduced. In addition, the detecting signal (detection flag) SD 1 of the voice detector VD1 is given to for example, voice switch VS, and is used for various control.

[0013] When the rate of the acoustic turnover component contained in the reference signal Vx is large according to the voice detector VD1 applied to this invention as mentioned above, while suspending processing of the background-noise power presumption section 3 by the background-noise power presumption change section 4 Since the judgment by the 1st voice / non-voice judging section 4 is performed based on the background-noise power estimate Pn which asked in the situation that the rate of the acoustic turnover component contained in the reference signal Vx is small, and the background-noise power presumption section 2 held, When the background noise in the telephone call terminal by the side of a far edge is sent out from a loudspeaker 11 and turns to a microphone 10 The situation that it becomes impossible to detect a sound signal even if the ratio of the voice component which the near end side speaker in the acoustic signal which a microphone 10 collects emits, and the other background-noise component originates in a bird clapper small and the sound signal is contained in the reference signal Vx can be reduced. Also in a system with a short distance of a before [ from a loudspeaker 11 / a microphone 10 ], and large acoustic turnover gain, it is detectable with a sufficient precision whether the reference signal Vx is a sound signal.

[0014] (Operation gestalt 2) Drawing 3 shows the block diagram of the voice detector VD2 in the operation gestalt 2 of this invention. However, since the fundamental composition of this operation gestalt is common in the operation gestalt 1, the same sign is given to common composition and explanation is omitted.

[0015] This operation gestalt is based on the detecting signal SD 1 of the 1st voice / non-voice judging section 3. the time tau1 and tau2 judged as the reference signal Vx being a non-voice signal as shown in drawing 4 , i.e., time for a detecting signal SD 1 to be L level, (henceforth "voice non-detecting duration"), and voice the non-detecting section which asks for -- a time check -- with the section 5 voice the non-detecting section -- a time check -- the feature is in the point equipped with the 2nd voice / non-voice judging section 6 which judges any of a sound signal and a non-voice signal the reference signals Vx are based on the voice non-detecting duration tau1 and tau2, and -- which were called for by the section 5

[0016] here -- voice the non-detecting section -- a time check -- whenever a detecting signal SD 1 changes from L to H in the section 5 -- a time check -- although processing is reset -- the time check before it -- between the phoneme duration grades in a sound signal holds the result (voice non-detecting duration tau1 --) for storage means, such as RAM, at least

[0017] The voice non-detecting duration tau1 and tau2 memorized by the storage means of the section 5, --, tauN are referred to. moreover -- the 2nd voice / non-voice section judging section 6 -- voice the non-detecting section -- a time check -- Over time for these values to be the phoneme duration grades of human being's voice, in being abbreviation regularity and being the pitch interval grade of human being's voice, these sections tau 1 - tauN are anew detected as the voice section, and it outputs the detecting signal SD 2 of H level (refer to drawing 4 ).

[0018].the ambient noise level [ in / the microphone 10 neighborhood / if it \*\* and twists in this operation gestalt, as it is shown in drawing 4 ] VN -- high -- the ratio of the instant power estimate PS and the background-noise power estimate Pn, since Ps/Pn is small In spite of containing voice in the reference signal Vx, when being judged as a non-voice signal in the 1st voice / non-voice judging section 3, it becomes possible to detect as a sound signal anew in the 2nd voice / non-voice judging section 6. Consequently, there is an advantage that a sound signal can be detected with a still more sufficient precision also in a \*\*\*\* telephone call system with large background-noise level to the operation gestalt 1.

[0019]

[Effect of the Invention] Invention of a claim 1 is used for the above-mentioned \*\*\*\* telephone call terminal of the \*\*\*\* telephone call system which the \*\*\*\* telephone call terminal which has a microphone and a loudspeaker is connected to other telephone call terminals or \*\*\*\* telephone call terminals, and performs a half-duplex telephone call. The instant power presumption section which presumes the instant power of the reference signal which is the voice detector which detects whether the signal transmitted to a channel is a sound signal, or it is a non-voice signal, and was taken out from the above-mentioned channel, The background-noise power presumption section which presumes the power of the background-noise component contained in the above-mentioned reference signal, While judging whether a reference signal is a sound signal or it is a non-voice signal based on the background-noise power estimate presumed in the instant power estimate and the above-mentioned background-noise power presumption section which are presumed in the above-mentioned instant power presumption section Since it had the 1st voice / non-voice judging section holding the last judgment result, and the background-noise power presumption change section which changes updating/halt of the background-noise power estimate in the above-mentioned background-noise power presumption section until the judgment result was updated When the rate of the acoustic turnover component contained in a reference signal is large, while suspending processing of the background-noise power presumption section by the background-noise power presumption change section Since the judgment by the 1st voice / non-voice judging section is performed based on the background-noise power estimate which asked in the situation that the rate of the acoustic turnover component contained in a reference signal is small, and the background-noise power presumption section held, When the background noise in the telephone call terminal by the side of a far edge is sent out from a loudspeaker and turns to a microphone The situation that it becomes impossible to detect a sound signal even if the ratio of the voice component which the near end side speaker in the acoustic signal which a microphone collects emits, and the other background-noise component originates in a bird clapper small and the sound signal is contained in the reference signal can be reduced. The effect that it is detectable with a sufficient precision is [ whether a reference signal is a sound signal and ] under the situation that background-noise level is large.

[0020] voice the non-detecting section where a reference signal finds voice the non-detecting duration detected as it is a non-voice signal according to the judgment result according [ invention of a claim 2 ] to the voice / non-voice judging section of the above 1st -- a time check -- with the section It has the 2nd voice / non-voice judging section which judges any of a sound signal and a non-voice signal the above-mentioned reference signals are from voice the non-detecting duration found by the section. this voice non-detecting section -- a time check -- Voice the non-detecting duration found by the section is abbreviation regularity over time to be the phoneme duration grade of human being's voice. this -- the 2nd voice / non-voice judging section -- the above-mentioned voice non-detecting section -- a time check -- And since all the reference signals of this voice non-detecting duration are judged to be a sound signal and it changes when the above-mentioned voice non-detecting duration is the pitch period grade of human being's voice [ when a sound signal is not detected in the 1st voice / non-voice judging section under the situation that background-noise level is very large ] voice the non-detecting section -- a time check -- almost uniformly, in being almost equal to the pitch interval of human being's voice, while voice the non-detecting duration measured in the section is an audio phoneme duration grade Since a reference signal is anew judged as a sound signal in the 2nd voice / non-voice judging section, it is effective in the ability to detect a sound signal with a still more sufficient precision in a \*\*\*\* telephone call system with large background-noise level.

---

[Translation done.]

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号  
特開2000-305579  
(P2000-305579A)

(43)公開日 平成12年11月2日 (2000.11.2)

(51)Int.Cl.  
G 10 L 11/02  
21/02

識別記号

F I  
G 10 L 3/00  
3/02  
9/00

5 1 3 A 5 D 0 1 5  
3 0 1 D  
D

マーク(参考)

審査請求 未請求 請求項の数2 O L (全 8 頁)

(21)出願番号

特願平11-115713

(22)出願日

平成11年4月23日 (1999.4.23)

(71)出願人 000005832

松下電工株式会社  
大阪府門真市大字門真1048番地

(72)発明者 福島 寛

大阪府門真市大字門真1048番地松下電工株  
式会社内

(72)発明者 竹山 博昭

大阪府門真市大字門真1048番地松下電工株  
式会社内

(74)代理人 100087767

弁理士 西川 恵清 (外1名)

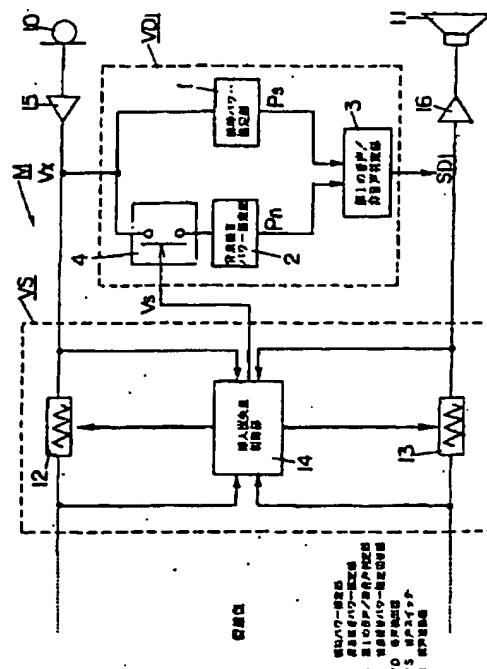
最終頁に続く

(54)【発明の名称】 音声検出器

(57)【要約】

【課題】音響結合利得が大きい系や背景雑音レベルが大きい状況下においても参照信号に音声信号が含まれているか否かを精度良く検出する。

【解決手段】背景雑音パワー推定切替部4では、参照信号Vxに含まれる音響結合成分の割合が大きいときに背景雑音パワー推定部2の推定処理を停止させる。これにより、背景雑音パワー推定値Pnが、通話状態によらずにマイクロホン10周辺の背景雑音パワーを近似した値となる。その結果、マイクロホン10とスピーカ11との間の距離が短く、音響結合利得が大きい系においても遠端側の背景雑音がマイクロホン10に回り込むことによる音声検出性能の劣化を低減し、背景雑音レベルが大きい状況下においても参照信号に音声信号が含まれているか否かを精度良く検出することができる。



## 【特許請求の範囲】

【請求項1】マイクロホン及びスピーカを有する拡声通話端末が他の通話端末又は拡声通話端末に接続されて半二重通話をを行う拡声通話系の上記拡声通話端末に用いられ、通話路に伝送される信号が音声信号であるか非音声信号であるかを検出する音声検出器であって、上記通話路から取り出した参照信号の瞬時パワーを推定する瞬時パワー推定部と、上記参照信号に含まれる背景雑音成分のパワーを推定する背景雑音パワー推定部と、上記瞬時パワー推定部で推定される瞬時パワー推定値並びに上記背景雑音パワー推定部で推定される背景雑音パワー推定値に基づいて参照信号が音声信号であるか非音声信号であるかを判定するとともに判定結果が更新されるまで前回の判定結果を保持する第1の音声／非音声判定部と、上記背景雑音パワー推定部における背景雑音パワー推定値の更新／停止を切り替える背景雑音パワー推定切替部とを備えたことを特徴とする音声検出器。

【請求項2】上記第1の音声／非音声判定部による判定結果に応じて参照信号が非音声信号であると検出された音声非検出継続時間を求める音声非検出区間計時部と、該音声非検出区間計時部により求められた音声非検出継続時間から上記参照信号が音声信号と非音声信号の何れであるかを判定する第2の音声／非音声判定部とを備え、該第2の音声／非音声判定部は、上記音声非検出区間計時部により求められる音声非検出継続時間が人間の音声の音韻継続時間程度の時間にわたって略一定であり、且つ上記音声非検出継続時間が人間の音声のピッチ周期程度である場合には該音声非検出継続時間の参照信号を全て音声信号と判定して成ることを特徴とする請求項1記載の音声検出器。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、住宅、事務所、工場等で用いられる拡声通話装置（インターホン、電話機、P H S 等）における通話回路に雑音除去機能や音声切替機能等を搭載するための音声検出器に関するものである。

## 【0002】

【従来の技術】一般に音声検出器は、マイクロホンにより集音された音響信号が音声信号を含んでいるか否かを検出するために用いられる。このような音声検出器の典型的な構成例を図1に示す。この音声検出器VD'は、瞬時パワー推定部20、背景雑音パワー推定部21並びに比較判定部22を備える。瞬時パワー推定部20は、立ち上がりが急峻であり且つ立ち下りが緩やかな特性をもつ積分回路又はデジタルフィルタ等により実現され、参照信号（マイクロホンにより集音される音響信号）の短時間平均パワーを推定するものである。また背景雑音パワー推定部21は、立ち上がりが緩やかであり且つ立ち下りが急峻な特性をもつ積分回路又はデジ

タルフィルタ等により実現され、参照信号中に定常的に存在する暗黙音（背景雑音）レベルを推定するものである。さらに比較判定部22は、瞬時パワー推定部20により求められる瞬時パワー推定値と、背景雑音パワー推定部21により求められる背景雑音パワー推定値の比を所定のしきい値と比較することにより、参照信号が音声信号を含んでいるか否かを判定（検出）してH又はLの2値信号（検出信号）を出力する。

## 【0003】

【発明が解決しようとする課題】ところが、スピーカ及びマイクロホンを有する拡声通話機（図示せず）の内部回路に上述のような音声検出器VD'を設ける場合、マイクロホンにより集音される音響信号にはスピーカからの回り込み成分（音響結合成分）が含まれる。この音響信号に含まれる回り込み成分の割合が大きい場合には、本来の目的であるマイクロホンの近傍に居る話者が音声を発したか否かを検出することが困難となる。例えば、遠端側の通話端末付近の背景雑音レベルが大きく、音響結合を経てマイクロホンが遠端側の背景雑音を集音する場合には、上記従来例における背景雑音パワー推定部21により求められる背景雑音パワー推定値が大きくなる。その結果、マイクロホンの近傍に居る話者が音声を発した状態においても、瞬時パワー推定値と背景雑音パワー推定値との比が小さく、所定のしきい値を越えることができず比較判定部22においては音声信号でない（非音声）として誤検出されてしまう虞がある。

【0004】本発明は上記問題に鑑みて為されたものであり、その目的とするところは、背景雑音レベルが大きい状況下においても参照信号に音声信号が含まれているか否かを精度良く検出することができる音声検出器を提供することにある。

## 【0005】

【課題を解決するための手段】請求項1の発明は、上記目的を達成するために、マイクロホン及びスピーカを有する拡声通話端末が他の通話端末又は拡声通話端末に接続されて半二重通話をを行う拡声通話系の上記拡声通話端末に用いられ、通話路に伝送される信号が音声信号であるか非音声信号であるかを検出する音声検出器であって、上記通話路から取り出した参照信号の瞬時パワーを推定する瞬時パワー推定部と、上記参照信号に含まれる背景雑音成分のパワーを推定する背景雑音パワー推定部と、上記瞬時パワー推定部で推定される瞬時パワー推定値並びに上記背景雑音パワー推定部で推定される背景雑音パワー推定値に基づいて参照信号が音声信号であるか非音声信号であるかを判定するとともに判定結果が更新されるまで前回の判定結果を保持する第1の音声／非音声判定部と、上記背景雑音パワー推定部における背景雑音パワー推定値の更新／停止を切り替える背景雑音パワー推定切替部とを備えたことを特徴とし、参照信号に含まれる音響結合成分の割合が大きい場合には背景雑音パ

ワーパー推定切替部によって背景雑音パワー推定部の処理を停止するとともに、参照信号に含まれる音響結合成分の割合が小さい状況で求められて背景雑音パワー推定部が保持していた背景雑音パワー推定値に基づいて第1の音声／非音声判定部による判定が行われるため、遠端側の通話端末における背景雑音がスピーカから送出されてマイクロホンに回り込むことにより、マイクロホンの集音する音響信号における近端側話者の発する音声成分とそれ以外の背景雑音成分との比が小さくなることに起因して参照信号中に音声信号が含まれていても音声信号が検出できなくなるような状況を低減することができ、背景雑音レベルが大きい状況下においても参照信号が音声信号であるか否かを精度良く検出することができる。

【0006】請求項2の発明は、請求項1の発明において、上記第1の音声／非音声判定部による判定結果に応じて参照信号が非音声信号であると検出された音声非検出継続時間を求める音声非検出区間計時部と、該音声非検出区間計時部により求められた音声非検出継続時間から上記参照信号が音声信号と非音声信号の何れであるかを判定する第2の音声／非音声判定部とを備え、該第2の音声／非音声判定部は、上記音声非検出区間計時部により求められる音声非検出継続時間が人間の音声の音韻継続時間程度の時間にわたって略一定であり、且つ上記音声非検出継続時間が人間の音声のピッチ周期程度である場合には該音声非検出継続時間の参照信号を全て音声信号と判定して成ることを特徴とし、背景雑音レベルが非常に大きい状況下で第1の音声／非音声判定部においては音声信号が検出されないような場合においても、音声非検出区間計時部において測定される音声非検出継続時間が音声の音韻継続時間程度の間ほぼ一定であり、且つ人間の音声のピッチ間隔とほぼ等しい場合には、第2の音声／非音声判定部において改めて参照信号を音声信号として判定するから、背景雑音レベルの大きい拡声通話系において音声信号をさらに精度良く検出することができる。

#### 【0007】

【発明の実施の形態】(実施形態1) 図1は、本発明の実施形態1における音声検出器VD<sub>1</sub>を有する拡声通話機Mを示すブロック図である。この拡声通話機Mは、マイクロホン10、スピーカ11、マイクロホンアンプ15、スピーカアンプ16、音声検出器VD<sub>1</sub>並びに音声スイッチVSを備え、回線を通じて他の拡声通話機等と接続される。ここで音声スイッチVSは、スピーカ11からマイクロホン10への音響結合、及び回線側での回り込みにより形成される閉ループの利得を低減させることによりハウリングを抑圧するものであり、マイクロホン10で集音する音声信号(送話信号)を回線へ伝送するための送話信号線上に挿入される送話側減衰器12と、回線から受信した音声信号(受話信号)をスピーカ11へ伝送するための受話信号線上に挿入される受話側

減衰器13と、通話状態に応じて送話側減衰器12並びに受話側減衰器13の利得を制御する挿入損失量制御部14とを備える。而して、挿入損失量制御部14においては、送受話信号を観測して通話状態を判定し、通話状態に応じて送話側減衰器12の利得及び受話側減衰器13の利得を適切に設定する。

【0008】本発明に係る音声検出器VD<sub>1</sub>は、通話路(送話信号線)から取り出した参照信号(送話信号)V<sub>s</sub>の瞬時パワーを推定する瞬時パワー推定部1と、参照信号V<sub>s</sub>に含まれる背景雑音成分のパワーを推定する背景雑音パワー推定部2と、瞬時パワー推定部1で推定される瞬時パワー推定値P<sub>s</sub>並びに背景雑音パワー推定部2で推定される背景雑音パワー推定値P<sub>n</sub>に基づいて参照信号V<sub>s</sub>が音声信号であるか非音声信号であるかを判定するとともに判定結果が更新されるまで前回の判定結果を保持する第1の音声／非音声判定部3と、背景雑音パワー推定部2における背景雑音パワー推定値P<sub>n</sub>の更新／停止を切り替える背景雑音パワー推定切替部4とを備える。

【0009】瞬時パワー推定部1は、立ち上がりが急峻であり、且つ立ち下がりが緩やかな特性をもつ積分回路又はデジタルフィルタ等によって構成される。また、背景雑音パワー推定部2は、立ち上がりが緩やかであり、且つ立ち下がりが急峻な特性をもつ積分回路又はデジタルフィルタによって構成される。

【0010】一方、第1の音声／非音声判定部3は、図2に示すように瞬時パワー推定部1から出力される瞬時パワー推定値P<sub>s</sub>を所定のしきい値P<sub>s0</sub>と比較してH又はLの2値信号D<sub>1</sub>を出力するコンバレータCP1と、瞬時パワー推定値P<sub>s</sub>と背景雑音パワー推定部2から出力される背景雑音パワー推定値P<sub>n</sub>との比P<sub>s</sub>/P<sub>n</sub>を求める除算器3aと、除算器3aの出力値P<sub>s</sub>/P<sub>n</sub>を所定のしきい値δと比較してH又はLの2値信号D<sub>2</sub>を出力するコンバレータCP2と、2つの2値信号D<sub>1</sub>、D<sub>2</sub>の論理積を求める論理積演算部3bとにより構成される。而して、本実施形態においては、瞬時パワー推定値P<sub>s</sub>がしきい値P<sub>s0</sub>よりも大きく(P<sub>s</sub>>P<sub>s0</sub>)、且つ除算器3aの出力P<sub>s</sub>/P<sub>s0</sub>がしきい値δよりも大きい(P<sub>s</sub>/P<sub>s0</sub>>δ)場合に音声信号と判定し、その他の場合に非音声信号と判定する。ここで、しきい値P<sub>s0</sub>は音声信号の最小レベルを規定するしきい値であり、しきい値δは音声信号レベルと背景騒音レベルとの最小比を規定するしきい値である。

【0011】背景雑音パワー推定切替部4は、音声スイッチVSの挿入損失量制御部14から出力される制御信号V<sub>s</sub>によりオンオフされて背景雑音パワー推定部2に対する参照信号V<sub>s</sub>の入力を切入するスイッチ等で構成される。そして、背景雑音パワー推定部2は、背景雑音パワー推定切替部4がオンして参照信号V<sub>s</sub>が入力されている場合に更新モードとなり、背景雑音パワー推定

切替部4がオフして参照信号Vxが入力されない場合に停止モードとなる。ここで更新モードにおいては、背景雑音パワー推定部2が参照信号Vxを参照して逐次背景雑音パワー推定値Pnを更新する。また、停止モードにおいては、背景雑音パワー推定部2が上記演算処理を停止し、背景雑音パワー推定値Pnとしてそれ以前に求められた値を保持する。

【0012】ここで、例えば音声スイッチVSの挿入損失量制御部14は、通話状態を受話状態と判定したときに制御信号Vsにより背景雑音パワー推定切替部4をオフするとともに、送話状態と判定したときに制御信号Vsにより背景雑音パワー推定切替部4をオンする。而して、背景雑音パワー推定部2では受話状態の時に停止モードとなり、送話状態の時に更新モードとなるから、音声検出器VD<sub>1</sub>においては、参照信号Vxに含まれる音響結合成分の割合が大きいときに背景雑音パワー推定部2の推定処理を停止することによって、背景雑音パワー推定値Pnを、通話状態によらずにマイクロホン10周辺の背景雑音パワーを近似した値とすることができます。その結果、マイクロホン10とスピーカ11との間の距離が短く、音響結合利得が大きい系においても遠端側の背景雑音がマイクロホン10に回り込むことによる音声検出性能の劣化を低減することができる。なお、音声検出器VD<sub>1</sub>の検出信号（検出フラグ）SD1は、例えば音声スイッチVSに与えられて種々の制御に利用される。

【0013】上述のように本発明に係る音声検出器VD<sub>1</sub>によれば、参照信号Vxに含まれる音響結合成分の割合が大きい場合には背景雑音パワー推定切替部4によって背景雑音パワー推定部3の処理を停止するとともに、参照信号Vxに含まれる音響結合成分の割合が小さい状況で求められて背景雑音パワー推定部2が保持していた背景雑音パワー推定値Pnに基づいて第1の音声／非音声判定部4による判定が行われるため、遠端側の通話端末における背景雑音がスピーカ11から送出されてマイクロホン10に回り込むことにより、マイクロホン10の集音する音響信号における近端側話者の発する音声成分とそれ以外の背景雑音成分との比が小さくなることに起因して参照信号Vx中に音声信号が含まれていても音声信号が検出できなくなるような状況を低減することができ、スピーカ11からマイクロホン10までの間の距離が短くて音響結合利得が大きい系においても参照信号Vxが音声信号であるか否かを精度良く検出することができる。

【0014】（実施形態2）図3は、本発明の実施形態2における音声検出器VD<sub>1</sub>のブロック図を示している。但し、本実施形態の基本的な構成は実施形態1と共通するので、共通する構成には同一の符号を付して説明を省略する。

【0015】本実施形態は、第1の音声／非音声判定部

3の検出信号SD1に基づいて、図4に示すように参照信号Vxが非音声信号であると判定された時間、すなわち検出信号SD1がLレベルである時間（以下、「音声非検出維続時間」という）t<sub>1</sub>、t<sub>2</sub>、…を求める音声非検出区間計時部5と、音声非検出区間計時部6により求められた音声非検出維続時間t<sub>1</sub>、t<sub>2</sub>、…に基づいて参照信号Vxが音声信号と非音声信号の何れであるかを判定する第2の音声／非音声判定部6とを備えた点に特徴がある。

【0016】ここで、音声非検出区間計時部6では、検出信号SD1がしからHに変化する毎に計時処理をリセットするが、それ以前の計時結果（音声非検出維続時間t<sub>1</sub>…）を少なくとも音声信号における音韻維続時間程度の間だけRAM等の記憶手段に保持している。

【0017】また第2の音声／非音声区間判定部6では、音声非検出区間計時部5の記憶手段に記憶された音声非検出維続時間t<sub>1</sub>、t<sub>2</sub>、…、t<sub>n</sub>を参照し、これらの値が人間の音声の音韻維続時間程度の時間にわたって略一定であり、且つ人間の音声のピッチ間隔程度である場合には、これらの区間t<sub>1</sub>～t<sub>n</sub>を改めて音声区間として検出し、Hレベルの検出信号SD2を出力する（図4参照）。

【0018】而して本実施形態に據れば、図4に示すようにマイクロホン10付近における周囲騒音レベルVnが高く、瞬時パワー推定値PSと背景雑音パワー推定値Pnとの比PS/Pnが小さいために、参照信号Vxに音声が含まれているにも拘わらず、第1の音声／非音声判定部4においては非音声信号として判定されてしまうような場合において、第2の音声／非音声判定部6において改めて音声信号として検出することが可能となる。その結果、実施形態1に対して背景雑音レベルの大きい拡声通話系においても音声信号をさらに精度良く検出することができるという利点がある。

【0019】

【発明の効果】請求項1の発明は、マイクロホン及びスピーカを有する拡声通話端末が他の通話端末又は拡声通話端末に接続されて半二重通話をを行う拡声通話系の上記拡声通話端末に用いられ、通話路に伝送される信号が音声信号であるか非音声信号であるかを検出する音声検出器であって、上記通話路から取り出した参照信号の瞬時パワーを推定する瞬時パワー推定部と、上記参照信号に含まれる背景雑音成分のパワーを推定する背景雑音パワー推定部と、上記瞬時パワー推定部で推定される瞬時パワー推定値並びに上記背景雑音パワー推定部で推定される背景雑音パワー推定値に基づいて参照信号が音声信号であるか非音声信号であるかを判定するとともに判定結果が更新されるまで前回の判定結果を保持する第1の音声／非音声判定部と、上記背景雑音パワー推定部における背景雑音パワー推定値の更新／停止を切り替える背景雑音パワー推定切替部とを備えたので、参照信号に含ま

れる音響結合成分の割合が大きい場合には背景雑音パワー推定切替部によって背景雑音パワー推定部の処理を停止するとともに、参照信号に含まれる音響結合成分の割合が小さい状況で求められて背景雑音パワー推定部が保持していた背景雑音パワー推定値に基づいて第1の音声／非音声判定部による判定が行われるため、遠端側の通話端末における背景雑音がスピーカから送出されてマイクロホンに回り込むことにより、マイクロホンの集音する音響信号における近端側話者の発する音声成分とそれ以外の背景雑音成分との比が小さくなることに起因して参照信号中に音声信号が含まれていても音声信号が検出できなくなるような状況を低減することができ、背景雑音レベルが大きい状況下においても参照信号が音声信号であるか否かを精度良く検出することができるという効果がある。

(0020) 請求項2の発明は、上記第1の音声／非音声判定部による判定結果に応じて参照信号が非音声信号であると検出された音声非検出継続時間を求める音声非検出区間計時部と、該音声非検出区間計時部により求められた音声非検出継続時間から上記参照信号が音声信号と非音声信号の何れであるかを判定する第2の音声／非音声判定部とを備え、該第2の音声／非音声判定部が、上記音声非検出区間計時部により求められる音声非検出継続時間が人間の音声の音韻継続時間程度の時間にわたって略一定であり、且つ上記音声非検出継続時間が人間の音声のピッチ周期程度である場合には該音声非検出継

続時間の参照信号を全て音声信号と判定して成るので、背景雑音レベルが非常に大きい状況下で第1の音声／非音声判定部においては音声信号が検出されないような場合においても、音声非検出区間計時部において測定される音声非検出継続時間が音声の音韻継続時間程度の間ほぼ一定であり、且つ人間の音声のピッチ間隔とほぼ等しい場合には、第2の音声／非音声判定部において改めて参照信号を音声信号として判定するから、背景雑音レベルの大きい拡声通話系において音声信号をさらに精度良く検出することができるという効果がある。

## 【図面の簡単な説明】

【図1】実施形態1を含む拡声通話機のブロック図である。

【図2】同上における第1の音声／非音声判定部のブロック図である。

【図3】実施形態2を示すブロック図である。

【図4】同上の動作説明用の信号波形図である。

【図5】従来例を示すブロック図である。

## 【符号の説明】

- 1 暫時パワー推定部
- 2 背景雑音パワー推定部
- 3 第1の音声／非音声判定部
- 4 背景雑音パワー推定切替部
- VD 音声換出器
- VS 音声スイッチ
- M 拡声通話機

【図2】

